Mathematics, Management, and Mind

Statement of Research Interests

Jon Doyle

Contents

Summary

My research interests center on artificial intelligence, mathematics, and the theory of computation, with secondary interests in philosophy, logic, economics, physics, history, science policy, and scientific publishing.

My most fundamental interests lie in helping develop a mathematical science of mind, and in developing architectures, representations, and methods for autonomous agents that reason and act in a rational, self-controlled manner. For the mathematical work, I study successful and interesting ideas to formalize hitherto unformalized notions (nonmonotonic logic provides the first example of this in my work), and periodically revisit existing formalizations to rethink them and to connect them with mathematical treatments of other notions. My current mathematical work addresses the topics of formalizing computationally useful notions of individual and group preference, relating these to nonmonotonic reasoning and notions of qualitative decision making, and developing theories of bounded rationality and rational reasoning. For the architectural work, I draw on the mathematical work and a ‘‘society of mind’’ metaphor that views architectures for individual agents closely related to architectures for collections of distributed interacting (cooperating, competing, or conflicting) agents. My recent work has addressed these problems in the contexts of medical informatics and military planning, as a participant in the MIT Guardian Angel project and in the DARPA/Rome Laboratory Planning and Scheduling Initiative.

I also have interests in social and governmental policy toward science and in the changes now affecting scientific publishing. I participated in writing a AAAI report to NSF, played major roles in the recent ACM/CRA Workshop on Strategic Directions in Computing Research, and address related issues as a AAAI Fellow and member of the AAAI Executive Council. I served as one of the initial associate editors of JAIR, one of the first electronic journals and the first electronic journal of artificial intelligence, continue to serve JAIR in an advisory capacity, and helped set up ACM Computing Surveys’ initial electronic publication offerings.

Toward a mathematical science of mind

To understand a physical phenomenon, current-day physics instructs us to isolate the system of interest, set up equations describing it, and solve the equations, either analytically or numerically, locally or globally, qualitatively or quantitatively, to determine the system's properties and possible behaviors. The known equations for describing systems and techniques for analyzing them support an array of vast and highly successful engineering disciplines (mechanical, electrical, chemical, and others) which use the knowledge of physics to predict and choose the properties of designs.

Physics did not begin this way, however. It began as natural philosophy, with no greater predictive power than that possessed by everyone. It grew into the mathematical science we know today through great observational labor and mathematical invention. For over two decades my main research aim has been to foster a similar evolution in thinking about mind, one that transforms current knowledge and study of mind from mental philosophy, in which people have a hard time telling if they are even talking about the same thing, into a mathematical science of mind that supports engineering disciplines capable of constructing reliable intellectual agents and appliances.

As with the transformation from seventeenth century natural philosophy to nineteenth century physics, the transformation of mental philosophy into mental science requires more than one person, and more than one decade. The initial stages of this effort necessarily appear somewhat chaotic as people try to make sense of myriad possibly mistaken or misemphasized observations. The early centuries of physical discussion concentrated on trying out various conflicting ways of looking at physical phenomena, ranging from the disputes between Heracliteans and Parmenideans about whether all is change or all is stasis to the seventeenth century interpretations of ‘‘force’’ as what we now call as mass, inertia, and other quantities. Rapid progress really began only after mathematical work by Newton, the Bernoullis, Euler, Cauchy and others made possible the fruitful conceptual identifications we know today, together with the mathematical laws connecting these concepts. This effort involved simultaneous mathematical invention and philosophical analysis, developing tools like differential calculus and identifying fruitful interpretations of mathematical constructs with physical phenomena or properties like linear momentum. It was a hard road for the pioneers, but we all enjoy the benefits of their struggles.

Current investigations of mental philosophy have some advantages over their natural philosophy antecedents, especially the example of the modern mathematical sciences, including modern mathematics and scientific methodology. The concepts and tools of logic, theory of computation, and mathematical economics and statistics certainly provide good initial help. These advantages do not fully overcome the central difficulties, however.

- As with earlier natural philosophy, mental philosophy exhibits a range of competing and conflicting views of the phenomena. Many of these lack adequate formalizations that might help compare and understand them. Rigorously developed ideas may bear up even though not tested through formalization, but the saying of Truesdell applies: the great advantage of mathematics over other fields is that mathematical statements can be wrong. To make good progress, we must be able to discard or reconsider some views as wrong rather than having to maintain them as ‘‘differing opinions’’. Moreover, mere formalization is not sufficient, as superficial logical encodings sometimes demonstrate. True insight often depends on deep mathematical analysis to disentangle the concepts involved in complex phenonmena.
- Worse still, the mathematical concepts needed for such formalizations may not yet exist; even where potentially applicable concepts do exist, the esoteric nature of some of them means few

people possess the knowledge or motivation to apply them to understanding mental phenomena. We certainly still lack much of the theoretical understanding that gives classical mathematics so much power in the physical realm, namely deep theorems that provide the ability to draw conclusions without numerical simulation. Current practice in AI relies on simulation, e.g., describing a system with a table of rules (playing the role of equations) and iteratively applying the rules to simulating the behavior of the system. Though simulation may be closer to reality in modeling reasoning than is numerical simulation in physics, since reasoning phenomena do seem to exhibit some of the step-by-step character of logical simulation, the lack of a deep predictive theory frustrates the theorist and impedes the practitioner.
- More fundamentally, the phenomena of interest appear qualitatively more complex than those analyzed in early natural philosophy: minds, reasoning, and decisions, rather than wheels, pulleys, and cannonballs. Early physics advanced in part on the backs of observational abilities to isolate different causes and conditions. Available methods for studying minds leave such isolation difficult, if not impossible, to achieve. In addition, common experience with the flexibility and subtlety (not to say deviousness and self-deception) of human minds makes it unlikely that small sets of experimental observations determine much about the subject of the observations without a strong theory backing them up.

In consequence, one must expect development of a science of mind to lean heavily on conceptual and mathematical experimentation, and to involve substantial work on developing new branches of mathematics. My own work toward this end started in the mid-1970s, but a growing number of people have become engaged in this activity over the years, and have made much good progress. Their efforts have definitely changed the landscape of AI.

My work toward a mathematical science of mind has focussed on taking specific problems and solutions and looking for ever deeper mathematical understanding of the concepts and techniques involved. For example, my early work on truth maintenance systems provided the key equilibrium and groundedness concepts for the initial nonmonotonic logic I developed with Drew McDermott. The logical formalization, though not perfect, served to open an entire new class of logical systems to technical investigation, and to significantly influence the subsequent theory of belief revision. This area has since grown into a large literature of nontrivial mathematical respectability and depth. My subsequent work reexamined the foundations of the logic, separated the mathematical essence from some superficial logical trappings, and identified connections with the theory of computation and economic notions of rationality.

Current projects

- Developing a useful qualitative decision theory, especially efficient qualitative representations of preferences and methods for using these preferences in decision making, either directly or through construction of quantitative multi-attribute utility measures. Preferences constitute a crucial element of intelligent agents, since virtually any display of intelligence involves some degree of judgment, and preferences provide the major element in capturing this judgment. The crude notion of preferences present in most software systems (i.e., selections from a set of options) cannot capture the economic notion of ordering relations or more complicated structures over finite or infinite sets of alternatives. The hard problems here concern both how to represent generic and defeasible preference information, how to infer preferences from plans and other information information about activities, and how to reason or make decisions with this information, either abstractly or through construction of quantitative measures.

- Finding comprehensive formalizations of limited rationality treating the interplay between reasoning and rationality, especially through interpretations of reasoned deliberation and nonmonotonic reasoning as embodying qualitative preferences and as guided by group decision-making rules. My earlier work on reasoned assumptions first drew the mathematical connections between sets of defeasible conclusions and economic equilibria of multi-agent choices. The continuing work seeks to use sets of reasons as primary descriptors of limits on reasoning abilities. Along the way I aim to complete a book in progress on the topic.
- Studying limits on rationality characterized in terms of mechanical notions of force and mass. This project, begun in the early 1980s, has served to inspire many of the successes in my other efforts.
- Developing a course and textbook of artificial intelligence from a mathematical point of view.
- Organizing expositions of outstanding mathematical problems in artificial intelligence.

Sample publications

- Non-monotonic logic I (with D. V. McDermott), *Artificial Intelligence* **13** (1980), 41-72.
- Impediments to universal preference-based default theories (with M. P. Wellman), *Artificial Intelligence* **49**(1-3) (May 1991), 97-128.
- A logic of relative desire (preliminary report) (with Y. Shoham and M. P. Wellman), *Methodologies for Intelligent Systems 6* (Z. W. Ras and M. Zemankova, eds.), Berlin: Springer-Verlag (1991), 16-31.
- Reasoned assumptions and rational psychology, *Fundamenta Informaticae*, **20**(1) (Spring 1994).

Architectures for rational self-management

Artificial intelligence has long sought to construct intelligent agents capable of undertaking activities on their own, either self-generated or in the service of human masters. For much of the time, however, work toward this end has proceeded without a clear model for the desired behavior, other than to point at how humans behave. Recent years have improved on this situation by identifying as target criteria some specific notions of rationality to be exhibited by artificial agents. Some of these notions of rationality draw on psychological notions of reasoning; some on logical notions of consistency and completeness; and some on economic ideals of rational choice. My work has sought to relate, combine, and understand all three of these conceptions of rationality, and to develop methods for making rational choices in reasoning, deliberation, planning, and representation, so as to construct what the Greeks called a *sophron*, a temperate agent making good choices in all activities.

Recent work has also sought to construct intelligent agents that collaborate with each other or with people. Indeed, the most natural context for computational agents is one of distributed action, of delegating responsibility for subtasks, cooperation with peers, and self-replication to pursue multiple ends. To avoid anarchy, these agents require means for coordinating their activities. More fundamentally, artificial agents themselves will not be unitary creations but instead will be composed of numerous sub-agents or faculties, with these sub-agents cooperating or competing in what Minsky and Papert call a ‘‘society of mind’’.

Current projects

- Exploring the use of market-guided control of problem-solving and self-management activities of an individual agent. I have been studying the use of market mechanisms to guide multiple agents,

partly as a means for coordinating among peers, and partly as one way of providing internal order in a society of mind, that is, as the framework for political economy among the faculties and subagents of a single mind. Preliminary results on methods for automatically constructing markets to reflect the structure of problem-solving goals suggest the utility of this approach, but much more work remains to be done, especially in modeling preferences about reasoning and computational activities in a reasonable way.
- Identifying and elaborating practical notions of reason (*né* truth) maintenance. The original notions impose severe limitations on the architecture of the reasoner, since the reason maintenance system must serve as the final authority about the current state of mind. I have been developing more realistic notions that abandon the unreasonable burden of authority in favor of functionality closer to what effective reasoners really need.
- Developing ‘‘ontologies’’ for and theories of plans and the entire planning process, from initial contemplation through postmortem analysis and archiving. These ontologies provide the backbone for construction of and communication among cooperating planning, scheduling, and workflow management processes.
- Developing Guardian Angels: personal, lifelong, active, medical assistants to help monitor one's health status, protect against healthcare errors, and explain ongoing and contemplated medical and lifestyle interventions. This project involves issues common to all intelligent agent efforts, together with work on representing medical knowledge and activities of patients and their doctors.

Sample publications

- A truth maintenance system, *Artificial Intelligence* **12** (1979), 231-272.
- A model for deliberation, action, and introspection, Ph.D. Dissertation, Massachusetts Institute of Technology, MIT/AI/TR-581 (May 1980).
- Guardian Angel: patient-centered health information systems (with P. Szolovits, W. J. Long, I. Kohane, and S. G. Pauker), MIT/LCS/TR-604 (May 1994).
- Toward rational planning and replanning: rational reason maintenance, reasoning economies, and qualitative preferences, *Advanced Planning Technology* (Austin Tate, editor), Menlo Park, California: AAAI Press, 1996, 130-135.

History, policy, and publishing

I have a long-standing interest in history, biography, and citation analysis. Even though the modern field of artificial intelligence is just over 40 years old, many of the important motivating ideas and many of the seminal discussions have been lost to recent students. In addition, most of the founders of the field still live. These circumstances present the need and opportunity for historical and biographical studies to illuminate the history of ideas in AI. Beyond writing personal recollections of some efforts in which I have been involved, I hope to help organize some of these studies, including collected works of some authors and long-term citation analyses.

Increasing competition for attention and funding has stimulated efforts by computer science and AI to clarify their aims, benefits, and importance to continued progress. I have been active recently in helping to summarize the state and future directions of computer science and AI, both as an aid to interesting students and others in pursuing the field, and as an aid to advocating governmental funding priorities, and expect to continue this work as needed. Related questions, especially the evolution of the structure of research and academic institutions, also interest me.

Interest in electronic publication has exploded recently, and some predict the demise of print journals within not too many years. This may well happen; in any event, on may expect to see important changes in the practice of science as a consequence of the new technologies. I have significant interests in the future of scholarly publication, and have been involved in a number of electronic publishing projects. I hope to help design structures and protocols that maximize the benefits to science and to scientists. Beyond the issues affecting day-to-day scientific activity, the solutions developed also need to keep an eye on history, lest excessive celebration of the new and transient cause trouble down the line for historians and scholarly-minded scientists.

Current projects

- A history of thinking about nonmonotonic reasoning prior to 1980.
- A brochure (and web site) to interest high-school students in AI.
- Development of a preprint and reviews archive for AI that also helps support conferences and journals.
- Finding ways of reducing the current proliferation of workshops and conferences that improve their value to science and scientists.

Sample publications

- A selected descriptor-indexed bibliography to the literature on belief revision (with P. London), *SIGART Newsletter*, No. 71 (1980), 7-23.
- Some Super-Classics of Artificial Intelligence? Privately circulated (1993), published on the web at http://www.medg.lcs.mit.edu/ftp/doyle/aiip93.ps
- Big problems for artificial intelligence *AI Magazine* **9**(1), 19-22 (1988).
- Strategic directions in artificial intelligence (with T. L. Dean), *ACM Computing Surveys* **28**(4) (December 1996).

Last modified: Thu Jan 9 09:23:05 EST 1997
Jon Doyle <doyle@mit.edu>